DOCUMENT RESUME

ED 223 412 SE 039 475

AUTHOR Myers, Jerome L.; And Others
TITLE Structure and Process in Learning Probability. Final Report.
INSTITUTION Massachusetts Univ., Amherst. Dept. of Psychology.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.; National Science Foundation, Washington, D.C.
PUB DATE 21 Oct 82
GRANT NIE-G-80-0126
NOTE 77p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *College Mathematics; *Educational Research; Higher Education; Instruction; *Instructional Materials; Mathematical Concepts; Mathematics Education; *Mathematics Instruction; *Probability; Problem Solving; *Textbook Research; Textbooks
IDENTIFIERS *Mathematics Education Research

ABSTRACT

Studies of methods of teaching statistics are reported. Three brief texts were written for the investigations; they varied in degree of explanation of basic concepts of elementary probability. Subjects were undergraduate students at the University of Massachusetts, Amherst. None of the subjects had had previous formal exposure to probability or statistics. All texts presented six basic formulas from elementary probability and an example of the application of each formula. All three documents contained definitions of special terms. The high-explanatory text emphasized the relation of probability to counting, used pictorial aids, and presented equations as natural developments ensuing from examples. The standard text lacked this conceptual development, but presented more concrete examples than those found in the low-explanatory document. Subjects were given sufficient time to read their text, then given an unrelated 15-minute task to minimize immediate memory dependence. This was followed first by a performance test, then by an aptitude measure. Results suggested that very different patterns appear to be present in the high-explanatory subjects as compared to the other groups. In particular, subjects in the high-explanatory group tended not to retrieve formulas, but to recall examples. (MP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED223412

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Final Report

Structure and Process in Learning Probability
NIE-G-80-0126

Principal Investigator: Jerome L. Myers
University of Massachusetts
Amherst, Massachusetts 01003
(413-545-2331)

Date: October 21, 1982

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
National Science Foundation
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SE039475

Final Report

Structure and Process in Learning Probability
NIE-G-80-0126

Principal Investigator: Jerome L. Myers
University of Massachusetts
Amherst, Massachusetts 01003
(413-545-2331)

Date: October 21, 1982

Elementary probability provides the foundation on which statistical theory and practice is built. Its importance as a discipline goes well beyond that, however. Frequently, the average citizen encounters probability concepts and statistics in his reading and in his viewing of television newscasts. To cite but a few examples, we are all exposed to information about statistical test results and correlation in discussions of the role of smoking in cancer, to information about confidence intervals in presentations of results of polls prior to elections, and even to probabilities in viewing sports and the weather. Many of these sorts of "everyday examples" are discussed in an excellent book of readings edited by Tanur (1978). We contend, then that statistical education is an important part of the general education of students, and therefore it is critical to attend closely to the fundamental concepts and calculations to be found in the area of elementary probability.

Despite the clear importance of the topic, there is no evidence that it is being taught properly. Indeed, research in the last decade (Tversky and Kahneman, 1974, provide a seminal paper) has made clear that college students have marked misconceptions about probability and statistical concepts. As Tversky and Kahneman note: "Misconceptions of chance are not limited to naive subjects." They go on to report results of studies conductd with research psychologists, individuals who ordinarily have had several statistics courses at the undergraduate and graduate level.

It appears time that systematic investigations of methods of teaching statistics were undertaken. The research conducted under the auspices of this grant provides a beginning. We begin with the observation that most introductory statistics texts emphasize memorization of formulas with little conceptual development. Recently, there have been a few exceptions to the general approach; Freedman, Pisani, and Purves (1978) and Pagano (1981) are examples of an approach that places a greater emphasis on understanding. We have conducted several studies contrasting these approaches and have obtained some interesting results. In what follows, I will summarize the general procedures and results, and provide a brief discussion. Three appendices to this report provide a considerably greater detail. The paper by Myers, Hansen, Robson, and McCann (Appendix 1) has been accepted for publication, that by Hansen, McCann, and Myers (Appendix 2) has been submitted for publication, and that by Myers (Appendix 3) constitutes an invited presentation to the First International Conference on the Teaching of Statistics and will appear in the published proceedings of that conference.

Methods

Subjects

Subjects in these studies were all undergraduate students at the University of Massachusetts, Amherst. None had previous formal exposure to probability or statistics.

Instructional Texts

Three brief texts were written for the purpose of this research program. During the course of the research program, revisions were developed to improve clarity of presentation. Such revisions were based on taped interviews with subjects who commented aloud as they read the text.

All three texts presented six basic formulas from elementary probability and an example of the application of each formula. All three texts contained definitions of such terms as independence and mutual exclusivity. In the high-explanatory text, the relation of probability to counting was emphasized, pictorial aids were used, and the equation was presented as a natural development ensuing from the example. The standard text lacked this conceptual development but presented examples that were somewhat more concrete than those employed in a low explanatory text. Further detail on the nature of the texts can be found in the Appendices.

Performance tests

In most of our studies these tests involved equal numbers of formula (e.g. "If P(A) = .4 and P(B) = .6, what is P(A and B)?") and story problems. There was one study (Appendix 2) in which only story problems were presented. Where both types of problems were presented, the order of presentation was counterbalanced. Examples of problems can be found in Appendix 2.

Procedure

Subjects were given sufficient time to read their text and then were given an interpolated unrelated task for 15 minutes to minimize dependence on immediate memory. They were then given sufficient time to do the performance test. Following this, they were given an aptitutde test based on items from the quantitative and analytical Graduate Record Examinations.

Results and Discussion

Very different patterns of knowledge appear to be present in subjects in the nonexplanatory (standard and low-explanatory texts) and explanatory conditions. Subjects in the two nonexplanatory groups performed considerably less well on story than on formula problems, and often used the correct formula for a problem (that is, met the lenient criterion) but failed to solve it. A closer look at answers to story problems revealed that subjects in the nonexplanatory conditions often required the explicit presence of key words which unambiguously pointed to certain operations, tended to misclassify problems in the presence of irrelevant or redundant information, and made many errors when the values in the story required modification before insertion into the formula. In contrast, subjects in the high-explanatory condition performed equally well on story and formula problems, tended to solve whenever they showed evidence of knowing the appropriate formula, and were considerably less hindered by absence of key words, the presence of irrelevant information, and the need to translate values in the story. The situation may be summarized by noting that, although all our subjects were novices, those in the high-explanatory condition seem to share more of the characteriscts of experts than do those in the other two conditions.

The results of more recent studies lead us to believe that the solution processes in the explanatory and nonexplanatory conditons are qualitatively different. In one study, subjects in a high-explanatory condition had only 60% correct recall of equations (as opposed to 90% for a standard group), but were able to perform correctly on 53% of story problesm. This result suggests to us (as does the equivalence of formula and story

scores in the study just detailed) that subjects in the high-explanatory condition often may not retrieve the formula at all. Another study in which subjects thought aloud while solving supports this conjecture; after studying a high-explanatory text, students tended to attempt to construct solutions, often using the diagrammatic aids provided by the text. Rather than recalling formulas, they attempted to recall examples from the text which they felt were similar to the problem at hand, and to use these as a model for solution.

Much remains to be done. High on the agenda must be the use of other measures of comprehension--both problems demanding more complex processing and transfer to new materials such as the Binomial and Bayes' theorems. We suspect that such research will provide further evidence of the role understanding can play in mathematical problem-solving.

References

Freedman, D., Pisani, R., and Purves, R. Statistics. New York: W. W. Norton, 1978.

Pagano, R. R. Understanding statistics in the behaviorial sciences. St. Paul: West, 1981.

Tanur, J. (editor). Statistics: A guide to the unknown. SanFrancisco: Holden-Day, 1978.

Tversky, A. and Kahneman, D. Judgement under uncertainty: Heuristics and biases. Science, 1974, 185, 1124-1131.

APPENDIX 1

The Role of Explanation in Learning

Elementary Probability

Jerome L. Myers, Randall S. Hansen,

Ricky C. Robson, and Joan McCann

University of Massachusetts, Amherst

Running Head: Learning Probability

Address: Jerome L. Myers

Department of Psychology

University of Massachusetts

Amherst, MA 01003

(Phone: 413-545-2653)

Abstract

In this study, 48 subjects who had no previous exposure to probability or statistics read one of three texts which varied in the degree of explanation of basic concepts of elementary probability. All texts contained six formulas, each accompanied by an example, as well as definitions and information logically required to solve all problems. The high-explanatory text differed from the low-explanatory and standard texts in that it emphasized the logical basis underlying the construction of the formulas, the relations among formulas, and the relations of variables to real world objects and events. On both immediate and delayed performance tests, subjects in the low-explanatory and standard text conditions performed better on formula than on story problems, while the subjects in the high-explanatory text condition did equally well on both types of problems. It was concluded that explanation did not improve the learning of formulas but rather facilitated the application of what was learned to story problems.

Parents, educators, and mathematicians have been involved in recurrent debates about the relative merits of understanding mathematical concepts and rote memorizing computational formulas. In principle, such conflicts should be resolvable through research involving manipulation of the degree to which students understand mathematical material and observation of their performance on tests of mathematical knowledge. In actuality, we encounter difficulties because we lack both a precise definition of mathematical understanding and a specification of the processes such understanding might affect.

The present study examines the effects of varying understanding of elementary probability. We have identified three dimensions of understanding and have constructed three texts that vary along these dimensions. The first dimension is similar to what Mayer and Greeno (1972) have labelled "external connectedness." In a high-explanatory text, probability was defined as a relative frequency of events, and examples emphasized the relation between answers and counts of the numbers of ways that various events could occur. A standard[1] text treated probability as a measure with certain well-defined properties (e.g. limits of one and zero) and used familiar real-world examples (e.g. the probability of rain tomorrow), but did not explain probability in terms of counting processes. A low-explanatory text was more abstract than either of the first two; aside from an initial reference to dice-throwing, examples were stated in terms of such quantities as P(A) and P(B).

The second dimension along which the texts varied was the degree of linkage among the equations presented. Only the high-explanatory text showed that one equation was a special case of another, or that one equation

could be derived from another. For example, this text pointed out that the probability of the union of two mutually exclusive events was a special case of the probability of the union of any two events.

The third dimension of relevance was the degree of explanation provided for each equation. In the high-explanatory text, Venn and branch diagrams were used to represent the information in the examples, and the equations applied in these examples were explained in terms of these diagrams. Neither of the other texts provided such explanations. For example, although all three texts stated that the probability of the joint occurrence of two independent events was the product of their probabilities and provided an example of this, only the high-explanatory text attempted to demonstrate why probabilities were multiplied in such cases.

Learning was measured by performance on two types of problems: formula problems, which merely presented certain values (e.g. "P(A) = .6, P(B) = .4") and necessary conditions ("A and B are independent events"), and required the subject to set up the correct results (e.g. "What is P(A and B)?" requires .6 x .4 as the answer); and story problems which were considerably less transparent. Solving formula problems should depend only upon memory of the equations, while solving story problems should depend also upon the ability to apply the equations; that is, to select the equation appropriate to the problem and then to insert the appropriate numbers from the story into the selected equation.

It is not clear which group should better remember the equations. There is some evidence that spending time and effort learning added, elaborative, material may hinder the acquisition of certain basic points (Reder and Anderson, 1980). In a like manner, the high-explanatory group

may spend proportionately less time and effort than the standard and low-explanatory groups on learning the equations; the high explanatory text involves 14 pages of material compared to four pages for the other two texts. On the other hand, subjects in the high-explanatory group should be able to relate each equation to both their world knowledge and the other equations; the required equations would therefore be more likely to be stored in memory, or stored with greater strength in memory. Such consequences of integration have been demonstrated by many investigators (e.g., Bransford & Johnson, 1972; Kintsch & van Dijk, 1978). Finally, the formulas may be easier to retrieve in the high-explanatory condition. When concepts are linked to many other concepts, there is a greater likelihood that the retrieval of any one piece of information will automatically "prime" the retrieval of other information (Collins & Loftus, 1975).

These arguments indicate that the high-explanatory group has potential advantages and disadvantages in learning and recalling the formulas. Thus, no text condition has an indisputable advantage on formula problems. Nor is it clear which group, if any, should perform best on story problems. Such problems may be viewed as involving three stages: categorization of the problem, retrieval of the formula appropriate for that category, and translation - correct substitution of values from the story into the retrieved equation. We would expect that because the high-explanatory text places an emphasis on understanding concepts and their connections to real-world referents, subjects in this condition would have an advantage in categorizing and translating story problems. As we noted above, however, the other texts may have an advantage in learning the formulas and, therefore, in retrieving them.

Although the potential advantage of the low-explanatory and standard groups in retrieving the formula makes it difficult to predict the outcome of comparisons between groups, there are two predictions that follow from the stage analysis just presented. First, because story problems involve categorization and translation, in addition to retrieval, they should prove considerably more difficult than formula problems. Second, this advantage of formula problems over story problems should be much less for the high-explanatory group than for the other two groups. In the high-explanatory text, external connectedness should aid in translating the components of a story problem; and linkage among equations and an understanding fostered by

the explanation of each equation may aid the subject in reconstructing the required equation.

This prediction of interaction finds some support in a study by Mayer (1974); he found that subjects had more difficulty solving story problems than formula problems and the effect was greater in a text emphasizing calculations (formula text). Mayer taught the binomial theorem using two texts which differed in their emphases on calculations and links to experience and in the sequencing of information. However, his general and formula texts are very different from any of our three texts in content, emphasis, sequencing of information, and length; and there are marked differences in our procedures and test problems. One potentially important difference is that Mayer's subjects had the binomial formula in front of them while taking the test so that the retrieval stage may not have been as critical as in the present study. Thus a different pattern of results is possible in the present study.

Parker (note 1) used materials more similar to those used in the present study. She tested eight subjects in each of two conditions, similar to our standard and high-explanatory conditions. She also found that story problems were more difficult than formula problems, and that this difference was more marked for her standard group. Following completion of that study, the high-explanatory text was modified to include questions designed to tap the subject's understanding of the text. This question-and-answer version was then read by eight new subjects, who thought aloud as they considered the interspersed questions. Their comments were tape-recorded and analyzed. On the basis of that analysis, several changes were made in the text used in this experiment. These changes were designed to clarify

the presentation. Subjects' responses on a questionnaire given at the end of the present experiment indicated that this goal was achieved.

In the present study, we have twice as many subjects in each condition, have added the low-explanatory text condition, have counterbalanced the order of posttests (which Parker failed to do), and asked subjects to set up the answer rather than to provide a final numerical result. This last procedural change allowed us to examine the possibility that subjects might use the correct formula but not substitute values correctly in it. This is important because our theoretical analysis suggests that such partially correct responses should be more prevalent in the standard and low-explanatory groups, which, presumably, have more difficulty with what we have called the translation stage.

Method

Subjects

Forty-eight volunteers from an Introductory Psychology course were randomly assigned to three text conditions. None had previous exposure to probability or statistics.

Text Materials

There were three texts: low-explanatory, standard, and high-explanatory. All began with a brief introduction indicating that the text contained a series of examples of probability calculations, and that each example would be followed by a formula which could be used to solve problems of that type. All three texts then contained a section labelled "What are the chances?" which introduced the concept of probability and its limiting values. This section noted that an action such as tossing a coin or observing the day's weather could have several possible outcomes, that a probability

value could be assigned to these outcomes, and that these values ranged from zero to one. The low-explanatory text then explained that the possible outcomes could be designated by letters, and introduced terms such as P(A) and P(B). The high-explanatory text was the only text to present probability as a relative frequency of outcomes. It presented a concrete example (a jar with six red and four white marbles) and explained that the probability of drawing a particular color was the proportion of marbles of that color.

The remainder of the three texts presented six examples and formulas. The formulas are presented in Table 1.

Insert Table 1 about here

The texts are too lengthy to reproduce here but some sense of the difference in approaches may be obtained by considering the first formula in Table 1. The low-explanatory text presented this material in the following way:

> "Two events, A and B, are said to be <u>incompatible</u> if only one of them can occur on a single trial. For example, a coin can come up either head or tail on a single toss so that these two events are incompatible. If A and B are incompatible, the probability that A or B occurs--P(A or B)--is the sum of the two probabilities. Suppose P(A) = .20 and P(B) = .45. Then,
>
> $$P(A \text{ or } B) = P(A) + P(B)$$
> $$= .20 + .45 = .65$$
>
> We might have several incompatible events which we'll call A, B, C, D, E, etc. Further suppose that we know the numerical values of P(A), P(B), and P(C). Then,
>
> $$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C) \qquad (1)$$

Out of a total of 100 marbles, we have 45 chances of drawing a red plus 20 chances of drawing a white one. Therefore, the probability of drawing a red or a white marble is

$$P(\text{red or white}) = \frac{45 + 20}{100}$$

$$= \frac{45}{100} + \frac{20}{100}$$

$$= P(\text{red}) + (\text{white})$$

We say that red and white are incompatible events because only one can occur on a single draw; the marble can be red, or white, but not both. A picture may help."

At this point, a figure containing a rectangle with two nonoverlapping circles was presented. One circle was labelled "45 red" and the other "20 white"; the area within the rectangle but outside the circles was labelled "35 other." Following this, the extension to more events was noted and Equation 1 was presented.

In summary, all texts contained an example followed by the equation, as well as a sufficient definition of mutual exclusivity. In the high-explanatory text, the relation of probability to counting was emphasized, pictorial aids were used, and the equation was presented as a natural development ensuing from the example.

Performance Tests

Two tests were constructed, each containing two sets of six story problems and two sets of six formula problems. Within each set, there was one problem testing each of the six equations in the text. The order of sets within a test was counterbalanced over subjects so that there were four sequences, with the constraint that story and formula sets would alternate. Half of

the subjects received one test first and the remaining subjects received the other test first.

Formula and story problems differed in the degree to which the given information was explicit in the question. For example, a formula problem testing knowledge of Equation 1 would state "P(A) = .4 and P(B) = .3. If A and B are incompatible, what is the value of P(A or B)?" In contrast, a story problem testing the same knowledge would read as follows: "The probability is .20 that Jimmy will win a race and .30 that he will finish second. What is the probability that he will win or finish second?"

Aptitude Tests

A ten-item test measured knowledge of basic algebra, decimals, percentages, fractions, and symbol translation. A second, eight-item test measured analytical ability; it required subjects to choose from among several Venn diagrams the one that represented a specified relationship among real-world sets. The items in both tests were taken from the analytical abilities and mathematics sections of the Graduate Record Examination aptitude test.

Procedure

Subjects were tested individually. They were instructed to read the text booklets silently and at their own rate, trying to understand the explanations and examples. The standard and low-explanatory groups, whose texts were four pages in length, were told that they would have 15 minutes to finish; the high-explanatory group, whose text was 14 pages in length, was given 25 minutes. These limits, based on pilot subjects, were imposed to discourage any individual from either quickly scanning the texts, or taking an inordinately long time. All subjects finished reading their texts in the given time period for their group. Subjects in the low-explanatory and standard groups

tended to rate the study time more favorably than subjects in the high-explanatory text condition, but this difference was not significant.[2]

At the completion of the study period, subjects were given two minutes to study a list of 15 words unrelated to probability and were then required to write down as many words as they could recall on a blank sheet of paper.

Upon completion of the interpolated task, subjects were given the test. There was one problem on a page with space to do any work which the subject felt might help in the solution. Subjects were given as much time as they required to complete the test; most required about 30 minutes. Upon completing the problems, subjects were asked to return 48 hours later. On the second test day, subjects were given the second form of the test. The order of presentation of the test forms was counterbalanced.

Following the second performance test, subjects were given as much time as they required to respond to the aptitude tests and to a questionnaire, which asked subjects to rate the study time, the comprehensibility of the text, and the difficulty of the problems. Since there were no significant differences between groups by any of these measures, they will not be considered further.

Results

Performance test data were scored according to both a strict and a lenient criterion. The strict criterion required that the subject set up the answer correctly in all respects; the lenient criterion gave credit if the subject used the correct equation but, for example, inserted the wrong values. This distinction applies only to performances on story problems; strict and lenient scores were nearly the same for formula problems.

Performance Test Data

Two analyses of variance were performed on the data. The first analysis was on the strict-criterion data. This involved three between-subject variables (three texts, four sequences of problems, and two orders of posttests) and three within-subject variables (two problem types, six sets of problems with one corresponding to each of the six equations, and two days). The second analysis dropped one problem type, the formula problems, and introduced criterion (strict versus lenient) as a variable.

Figure 1 summarizes the results. Consider the analysis of the strict-

---

Insert Figure 1 about here

---

criterion data first. Formula problems had a higher rate of solution than story problems when scores were averaged over text conditions and days; $F(1,24) = 532.22$, $p < .001$, $MSE = .6129$. Problem type interacted with text; $F(2,24) = 10.41$, $p < .001$, $MSE = .6129$. Two subsidiary analyses were performed to determine the source of this interaction. First, simple effects tests were performed comparing the three text conditions on formula and on story scores; despite the apparent superiority of the standard and low-explanatory groups on formula problems, and of the high-explanatory group on story problems, neither analysis yielded a significant result ($p > .10$). Analyses of covariance using aptitude scores as covariates produced the same result.

In a second analysis, story and formula scores were contrasted for each of the three text conditions. The low-explanatory and standard groups did much better on formula than on story problems while the high-explanatory group performed about equally well on both problem types. F tests of problem

type were performed on the data for each group separately: for the low-explanatory group, $F(1,24) = 33.66$, $p < .001$; for the standard group, $F(1,24) = 39.16$, $p < .001$; for the high explanatory group, $F < 1$. Because the error variances were quite stable across the three text conditions, the pooled MSE of .6129 was the error term in all three tests.

The effect of days differed for the two types of problems, as indicated by the significant day x problem type interaction; $F(1,24) = 7.12$, MSE = .1580, $p < .02$. There was a small nonsignificant drop in proportion correct for the formula problems from the first to the second test (the proportions are .63 and .60, respectively), and a small nonsignificant increase for the story problems (.43 and .46).

Table 2 presents the proportion correct for each group for each set of four problems testing knowledge of each equation. The problem sets varied

---

Insert Table 2 about here

---

in difficulty [$F(5,120) = 21.72$, MSE = 5.63, $p < .001$] ; the degree of variability among the six sets depended upon text condition [$F(10,120) = 2.09$, MSE = .5163, $p < .05$] and problem type [$F(5,120) = 18.12$, MSE = .3566, $p < .001$]. One factor in this variability appears to be that mean scores for the standard and low-explanatory groups are well below those of the high-explanatory group on all the story problem sets testing knowledge of Equations 1, 2, 4, and 5 (Set A) but not on those testing knowledge of Equations 3 and 6 (Set B). In response to this observation, four significance tests were performed contrasting the high explanatory group against the other two groups combined on Set A story problems, on Set B story problems;

on Set A formula problems, and on Set B formula problems. Only the F ratio for Set A story problems was greater than 1; $F(1,45) = 6.19$, $p < .025$.

This contrast is post hoc and the reported F statistic should be viewed with caution; it would not be significant using Scheffé's criterion. Nevertheless, the result makes sense in terms of the stage model presented earlier. Story problems testing knowledge of Equations 3 and 6 (Set B) contain a key word, "and," which unambiguously dictates multiplication of the stated probabilities. This is not true of other story problems; for example, "or" may dictate Equation 1 (addition) or Equation 2 (addition and subtraction). This difference between Set A and Set B story problems suggests that the difference in results in testing contrasts on the two sets reflects an inability of subjects in nonexplanatory conditions to match the performance of those in the high-explanatory group except when categoriaztion and translation requirements are minimal.

The analyses reported so far have been based upon a strict scoring criterion; to receive credit for an answer, the subject had to have it entirely correct. Answers to story problems were also scored according to a lenient criterion, which accepted as correct the appropriate formula, regardless of whether values were correctly inserted. The criterion did not interact with the sequence of test problems within or across days (all Fs < 1) and, therefore, these sources were pooled with the error term for criterion and criterion by text condition to permit more powerful tests. Lenient criterion scores were significantly higher than strict criterion scores; $F(1,45) = 49.59$, $MSE = .3656$, $p < .01$. The criterion interacted significantly with text condition, $F(2,45) = 3.65$, $p < .05$. The reason for this interaction appears to be that strict and lenient criterion scores differed much less in the

high-explanatory text condition than in the other two conditions; the interaction sum of squares was completely accounted for by the contrast between the high-explanatory and the other two conditions with respect to the strict-lenient difference. The F ratio for this contrast was 7.29, $p < .05$ using Scheffe's procedure.

Aptitude Data

The respective mean aptitude scores for the low-explanatory, standard, and high-explanatory groups were the following (expressed as proportion correct): Quantitative - .66, .66, .78; Analytical - .69, .71, .82. The differences were not significant, both ps > .10. Furthermore, an additional analysis of performance scores for a subset of subjects matched in aptitude produced the same results as the original analysis reported in the preceding sections.

Discussion

Consider three alternative models of the role of explanatory text. An elaboration model holds that the presence of additional, integrative and explanatory, material provides additional retrieval routes to the main points to be remembered (Reder & Anderson, 1980). That model provides an account of many memory phenomena, including levels of processing results, and better memory for superordinate ideas in passages (Anderson, 1976). It would predict that the high-explanatory group would have an advantage in retrieving formulas. It clearly does not provide an adequate account of the effects of our texts. Our high-explanatory group does worse than the others (although not significantly) in solving formula problems. Furthermore, in a study recently completed, we have found significant superiority of the standard group against the high-explanatory group (.90 vs. .60 correct)

in recalling the right hand side of the equations to the cue of the left hand side.

While the elaboration model focuses on retrieval, the interference model focuses on initial storage. It holds that the additional material inherent in an elaborated text takes needed resources and the main points are less well learned. Reder and Anderson (1980) concluded that this is what happens in studying chapters from texts in linguistics and geography. The results on formula problems and memory for formulas, noted above, suggest that there may be some truth to this. Formulas do appear to be less well stored in the high-explanatory condition. Nevertheless, this model is incomplete as a characterization of mathematical learning because it fails to come to grips with the processing of story problems.

The multistage model presented earlier in this paper provides a better account of our data. Within this framework, the results of the present study can be explained by two assumptions; the first is that both categorization and translation are more difficult for story than for formula problems; in story problems it is less obvious what kind of problem is present and, once that is decided and an equation retrieved, it is less obvious which numbers go where. The second assumption is that external connectedness, linkage among equations, and explanations of equations--qualities primarily present in the elaborate high-explanatory text--in some combination facilitate categorization and translation.

From these assumptions, it follows directly that subjects in the high-explanatory group should be less affected by problem type than those in the standard and low-explanatory groups, should be more likely to solve problems when they had retrieved the appropriate equation and should be aided less by the presence of key words in selecting appropriate equations

for story problems. All three of these consequences of the model were supported by the data. First, while the standard and low-explanatory groups performed significantly less well on story than on formula problems, the high-explanatory group performed about the same on both. Second, lenient and strict criterion scores were further apart for the low-explanatory and standard groups than for the high-explanatory group, supporting the assumption that translation of the story was more difficult for the first two groups. Third, the two nonexplanatory groups performed as well as the high-explanatory group only on story problems in which the wording unambiguously requires a particular operation. The advantage of the high-explanatory group is much larger for other story problems. An examination of the kinds of errors subjects made on individual problems provided additional evidence about the effects of text. The performance of the high-explanatory group was less affected by the presence or absence of key words, and was less depressed by the inclusion of irrelevant information in the statement of story problems. Since this analysis is post hoc and unsupported by significance tests, the support it offers our model is at best tenuous. It does suggest, however, that future studies might profitably manipulate the presence of key words and irrelevant information, and study the interaction with text condition.

The greater reliance of the low-explanatory and standard groups on key words parallels a result reported by Chi, Feltovich, and Glaser (note 2) who found that novices learning physics were more dependent upon such surface features than were experts. They note that both groups may use the same set of key words but "the actual cues used by the experts are not the words themselves but what they signify." In our study, such words may be immediate retrieval cues for the low-explanatory and standard groups but may signify secondary cues such as intersections and unions for the high-

explanatory subjects. The present work complements studies comparing experts and novices. Although all our subjects were novices, those taught with the high-explanatory text differed from other subjects in their processing of story problems in ways similar to those in which experts have been shown to differ from novices. If the goal of instruction is to move the novice along the path toward expertise, the role of explanation would seem to be critical.

Reference Notes

1. Parker, J. Understanding probability, Senior Honors Thesis, University of Massachusetts: Department of Psychology, May 1980.

2. Chi, M. T. H., Feltovich, P. J., & Glaser, R. Categorization and representation of physics problems by experts and novices (Tech. Rep. 4). University of Pittsburgh: Learning Research and Development Center, May 1981.

References

Anderson, J. R. Language, memory, and thought. Hillsdale, N. J.: Erlbaum Associates, 1976.

Bransford, J. D. & Johnson, M. K. Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 717-721.

Collins, A. M. & Loftus, E. F. A spreading-activation theory of semantic processing. Psychological Review, 1975, 82, 407-428.

Kintsch, W. and van Dijk, T. A. Toward a model of text comprehension and production. Psychological Review, 1978, 85, 363-394.

Mayer, R. E. Acquisition processes and resilience under varying testing conditions for structurally different problem-solving procedures. Journal of Educational Psychology, 1974, 66, 644-656.

Mayer, R. E. & Greeno, J. G. Structural differences between learning outcomes produced by different instructional methods. Journal of Educational Psychology, 1972, 63, 165-173.

Mayer, R. E., Stiehl, C. C., & Greeno, J. G. Acquisition of understanding and skill in relation to subjects' preparation and meaningfulness of instruction. Journal of Educational Psychology, 1975, 67, 331-350.

Reder, L. M. & Anderson, J. R. A comparison of texts and their summaries: Memorial consequences. Journal of Verbal Learning and Verbal Behavior, 1980, 19, 121-134.

Table 1

Formulas Presented in the Three Texts

(1) P(A or B or C) = P(A) + P(B) + P(C)

(A, B, and C are mutually exclusive)

(2) P(A or B) = P(A) + P(B) - P(A and B)

(general case)

(3) P(A and B and C) = P(A) x P(B) x P(C)

(A, B, and C are independent)

(4) P(A, B, B, B, A) = P(A) x P(B) x P(B) x P(B) x P(A)

(5) P(A given B) = P(A and B)/P(B)

(6) P(A and B) = P(A given B) x P(B)

Table 2

Proportion of Correct Responses for Each Equation For Story and Formula Problems, Pooled over Days

Story Data

Group	Equation 1	2	3	4	5	6	Mean
Low-Explanatory	.531	.172	.547	.406	.203	.594	.409
Standard	.547	.219	.531	.250	.219	.609	.396
High-Explanatory	.828	.297	.563	.531	.359	.625	.534
Mean	.635	.229	.547	.396	.260	.609	.446

Formula Data

Group	Equation 1	2	3	4	5	6	Mean
Low-Explanatory	.750	.484	.813	.750	.594	.453	.641
Standard	.578	.578	.781	.766	.719	.453	.646
High-Explanatory	.750	.391	.594	.672	.422	.484	.552
Mean	.693	.484	.735	.729	.578	.463	.613

APPENDIX 2

Students' Problem Solving Strategies and Patterns of Errors in Learning Elementary Probability

Randall S. Hansen, Joan McCann,

and Jerome L. Myers

Running Head: Learning Probability

Address: Jerome L. Myers

Department of Psychology

University of Massachusetts

Amherst, MA 01003

(Phone: 413-545-2331)

Statistical inference is part of the research methodology of many disciplines. It is not surprising, therefore, to find introductory statistics courses required, or highly recommended, in the undergraduate curricula of such diverse units as departments in the social and biological sciences, and in schools of education, business, public health, and agriculture. A basic element in such courses is probability: typically, the addition and multiplication theorems, the binomial theorem, and possibly conditional and joint probabilities. Given the importance of statistics in general, and probability in particular, an understanding of the consequences of various approaches to teaching this material is clearly desirable. The need for research on the teaching of probability and statistics becomes even clearer when we note the considerable evidence that most adults do not understand probability and, in fact, have pronounced misconceptions (Tversky and Kahneman, 1974).

In the last decade, several investigators (Mayer, 1974; Mayer and Greeno, 1972; Mayer, Stiehl, and Greeno, 1975; Myers, Hansen, Robson, and McCann, 1982) have varied the relative emphasis placed upon rote and conceptual learning in teaching elementary probability. The most consistent finding in all of these studies is that neither methods which emphasized rote learning of formulas nor those which emphasized understanding of concepts provided a uniform advantage of post-instruction tests. Performance was a function of both the method of instruction and the type of problems.

These results are based on group means for sets of problems. Because of this, they provide only limited insights into the processes involved when novices attempt to solve elementary probability problems. To further

our understanding of those processes, we have recorded clinical interviews with individuals attempting to solve problems after brief exposure to a text which presented, and attempted to explain, a few basic probability formulas. Some of the results of those interviews will be presented and discussed in the next section. We have also returned to the data originally collected by Myers, et al., (1982) to perform a more molecular analysis. More precisely, we have examined the types of errors made on individual problems by subjects taught by texts which differed with respect to the relative weight placed upon rote learning and understanding of probability formulas. Those error analyses also will be considered in this article.

PROTOCOL ANALYSES

Method

Subjects. Nine undergraduate students at the University of Massachusetts each received six dollars for their participation. None had previous exposure to probability or statistics.

Text material. The text was 14 pages in length, and contained concepts and equations pertaining to elementary probability. It explained such terms as mutual exclusivity and independence, and presented six equations, each of which was prefaced by an example. The equations are shown in Table 1. Probability was defined as a relative frequency of events, and examples emphasized the relation between answers and counts of the number of ways that various events could occur. Equations were developed in terms of their relation to other equations; for example, the text noted that the probability of the union of two mutually exclusive events (Equation 1) was a special case of the probability of the union of any two events (Equation

2). The text used pictorial aids, such as Venn and branch diagrams, to explain why an equation was applied in a particular example.

Insert Table 1 about here

Procedure. Subjects were tested individually. They were instructed to read the text at their own rate, trying to understand the explanations and examples. They were permitted to take notes, but were informed that they would not be allowed to use them during the subsequent problem solving period. Most subjects required about 30 minutes to read the text. At the conclusion of the study period, subjects were given two minutes to study a list of 15 words unrelated to probability and were then asked to write down as many words as they could recall on a blank sheet of paper. Upon completion of the interpolated task, subjects participated in a tape-recorded interview in which they were instructed to "think out loud" as they attempted to solve a series of 16 problems related to the text. A typical interview lasted approximately 1 hour. (Three subejcts who required that much time for their first few problems were not given all 16.)

When subjects first considered a problem, the interviewer was non-directive, allowing subjects to freely answer and explain their reasoning. If subjects encoutered extreme difficulty with a problem, the interviewer provided direction by such means as questioning certain aspects of the subject's reasoning and encouraging the subject to pictorially represent the information in the problem. The interviewer also reminded subjects to think aloud when they fell silent.

Problems. The problem set consisted of two story problems for each of the six equations in the text, as well as four "conceptual" problems which were included as an alternative means of taping subjects' understanding of the text. The solution of these last four problems required subjects to reason beyond the information provided by a formula.

Results and Discussion

Before considering the problem solving processes, as inferred from protocols, it may be useful to have a sense of the level of performance. Answers were scored correct if correctly set up in all respects. Subjects were not required to calculate the numerical result. The proportion correct for the six subjects who completed the problem set was .90 on story problems and .79 on conceptual problems. The three remaining subjects attempted an average of 5.33 story problems with a mean proportion correct of .56.

In examining the interview protocols, we focussed on two questions: How did subjects attempt to categorize problems? How helpful were the diagrammatic aids provided in the text? We now turn to these matters.

Categorization of problems. Categorization is a crucial stage in problem solving (Hinsley, Hays & Simon, 1977). Cognitive psychologists have suggested a number of ways in which instances of a category may be classified, among them categorization on the basis of critical features (Bourne, Dominowski, and Loftus, 1979) and categorization by analogy to previously experienced instances of the category (Brooks, 1978). Our subjects exhibited both of these processes. We will consider each in turn.

Subjects frequently used key words such as "and," "or," or "given" as cues to categorization of a problem. Following are two typical protocols. The first demonstrates how a subject decides between addition and multiplication:

S:...If I was relating different events by an "and" instead of an "or," then I would multiply the probabilities, whereas if they were "or" - then I would add the probabilities. That's the way I looked at it while I was doing it. So I'm trying to see whether these are "ands" or "ors."

Other protocols illustrate dependence on the key word "given." Note that the subject in the following excerpt is somewhat confused about exactly what goes into the numerator of the ratio of probabilities:

S:...I remember at the end-there was a formula that showed you how to calculate someting given something else. And it had to do with a ratio--a ratio of these two--added together--...A given B--I remember that part--was equal to the ratio where the numerator was A plus B.

Reliance on key words is only partially successful. Even if the problem is categorized correctly, the key word strategy provides no help in substituting the correct values from the problem into the equation. The key word strategy often does not unambiguously distinguish between equations (for example, between applications of Equations 1 and 2); and it is inapplicable when the key word is not explicit in the text. We shall see evidence of these difficulties in the section on error analyses.

Finally, even when applicable, the key word strategy is subject to misuse. For example, protocols demonstrated frequent confusion about which word signified multiplication and which addition. We shall see evidence of these difficulties in the section on error analyses.

Although key words were clearly important to subjects, they also occasionally verbalized their recognition of the similarity of a problem to

an example seen in the text. Following is an excerpt from a protocol of a subject who, in trying to solve a problem, notes that he had seen an analogous example in the text:

Subject(S): Okay, 50% is under 40, 60% is male, 20% is both... This one seems to be like that ace of hearts one (an example from the text)[1].

Instructor(I): Why do you say that?

S: Because, it overlaps-there's part of it that applies to both 40 and male. So under 40 or male--I know I subtract something--

I: What do you think you should subtract?

S: I subtract this.

I: The 20% under 40 and male?

S: Uh-huh

I: Why do you think you should subtract that?

S: Because I remember the one--you were supposed to subtract the ace of hearts--that was the overlap. So I think these involve addition.

I: Why do you think that?

S: Just the way I remember it in my mind--the problem was laid out in the text...It was the chance of drawing an ace or a heart and they got three chances of drawing an ace, 13 of drawing hearts and it was minus one for the ace of hearts. So--it doesn't come out to the same answer that they gave.

I: Do you know why?

S: I think it came out to 17. Oh, maybe I should put four aces.

I: Does that make sense to you--4 aces and 13 hearts?

S: Yeah, 'cause that's the whole thing. And that's just minus the overlap. So now I'll go up here and do this up here. Four aces, thirteen hearts, 50% under 40, 60% male, 20% under 40 and male; so that would be .5 plus 60% male minus .2 equal to .9. That's a lot.

Protocols such as this demonstrate that subjects were able to go beyond superficial features and extract classification cues. The subject in the example noticed the intersection in the problem, used this cue to recall an example, and analyzed the example to arrive at the solution. In short, the subject made use of conceptual relations between the problem and the example, rather than relying on memorized associations between surface features of the problem statement (e.g., the word "or") and a formula.

The role of diagrams in the text. Venn and branch diagrams were included with the presentation of each equation in the text in order to aid subjects in understanding the rationales for these equations, and to provide a framework within which to solve problems. Comments by subjects in this study and in a previous one (in which subjects were asked to think aloud as they worked through a question-and-answer version of the text) indicate that the diagrams did help subjects understand the text. However, they were less successfully used in solving test problems. Although most subjects attempted to use at least one of these representations at some point during the problem solving session, they did not attempt to use them very frequently.

Mistakes involving Venn diagrams were typically due to confusion about how to construct and label the diagram, rather than in inability to interpret a correct representation. For example, one subject made three attempts to represent the information given in the problem presented in footnote 1. In the first two attempts, the subject was unable to represent the information that 60% of the city's population was male. This was because he focussed on the probability of being under forty and its complement, first dividing a single circle into areas of "over" and "under 40," and then drawing a diagram containing an intersection of "over and under 40." In the third attempt at a representation, the subject drew two non-overlapping circles for "over" and "under 40," with a third circle, labelled "males," overlapping each other. This diagram, however, contained an area which supposedly represented males that are not over or under 40.

Misrepresentation involving branch diagrams seemed to occur most often because subjects lacked the ability to move from a representation involving relatively few concrete, countable events (as given in the text), to one that involved more abstract probabilities of events that were more difficult to count. While the subject in the following example did not actually draw a branch diagram, the protocol is relevant in that it shows the subject attempting to obtain the same result that a branch diagram would yield; namely, the number of events in the sample space and the number of events of interest. The following is an excerpt of a subject attempting to solve the following problem:

> "On a multiple choice test with 4 choices for each item, a student guesses. What is the probability that he is right on items 1 and 2 but wrong on 3 and 4?"

Subject(S): Probability of being right is one fourth, .25. And wrong would be .75.

Interviewer(I): Okay, how did you get those two?

S: Well, out of the four choices for each item, you get to choose one--so that would be one out of four chances--you have one chance of being right--which would be one to four which would be .25. And to be wrong--there are three wrong answers and one right answer, so out of four questions with four choices each--there's sixteen possible choices--out of sixteen possible choices, four could be right, twelve could be wrong. Chances of getting them all right would be .25.

I: The chances of getting the first one right, the second one right, and the third and fourth is .25?

S: Yeah.

I: Okay, why do you say that?

S: Because, out of the sixteen possible choices for all four questions, there are only four right answers, so it would be four sixteenths--one fourth--I'm going to take a wild guess at this one--and say it's one eighth.

I: Why do you say that?

S: Because to get them all right is--it's .25. That's the chance you get, one quarter chance. So you're halving--wait a second--you's halving that.

I: Why are you halving that?

S: Well, you're only getting two right--but you're only getting two wrong, too,--The chances are better for getting things wrong--I don't know.

This subject has made a fairly reasonable attempt to obtain the answer as a relative frequency. The major error occurred in defining an event as one of sixteen alternative answers (four for each question), instead as one of 256 possible sequences of answers. (The text had not explained that probabilities may be assigned to segments of a branch diagram in order to simplify it. When all segments are equally weighted, the correct diagram has 256 possible paths, nine of which satisfy the conditions specified in the problem.) This subject seems to have been aware that her representation was incorrect, that it failed to compensate for the greater probability of guessing incorrectly. However, she was unable to use this knowledge correctly. The subject seems to have been prepared to think in terms of a fairly small, countable sample space, but unable to extend this thinking to situations in which very large sample spaces or abstract probabilities are involved. Although the use of branch diagrams by our subjects was infrequent, this difficulty was typical on those occasions when they were employed.

ANALYSES OF ERRORS

In the study just described, we employed only one text. An earlier study by Myers et al., (1982) provides a basis for comparing the effects of instruction by that text (which we will call "explanatory") with the effects of instruction by texts which provide little explanation, or integration, of the equations taught. In this section, we provide a more molecular

analysis of the data than is available in the earlier article. By examining the kinds of errors made on individual problems under different text conditions, we arrive at some sense of the way in which different instructional modes influence what is learned and how it is used in solving probability problems.

Method

Subjects. Forty-eight volunteers from an Introductory Psychology course were randomly assigned to one of three text conditions. None had previous exposure to instruction in probability or statistics.

Texts. There wre three text conditions: The explanatory text was the same as the text used in the protocol study described in the preceding section. A second text was labeled "standard" because it seemed similar to many introductory statistics textbooks currently in use; it treated probability as a measure with certain well defined properties (e.g., limits of one and zero) and used familiar real-world examples (e.g., the probability of rain tomorrow), but did not explain probability in terms of counting processes. Nor did the standard text provide any explanation or develop relations between equations. A low-explanatory text was even more abstract than the standard text; aside from an initial reference to dice throwing, examples were stated in terms of such quantities as P(A) and P(B). All three texts presented an example followed by the appropriate equation, as well as definitions of terms such as independence and mutual exclusivity. A more detailed presentation of the material in the three texts may be found in Myers et al., (1982).

Problems. The 24 story problems included the 12 used in the protocol analysis study and 12 others that were similar. There were four problems testing knowledge of each of the six equations presented in the texts.

Procedure. Subjects were tested individually. They were instructed to read the text booklets silently and at their own rate, trying to understand the explanations and examples. The standard and low-explanatory groups, whose texts were four pages in length, were told that they would have 14 minutes to finish; the explanatory group, whose text was 14 pages in length, were given 25 minutes. These limits, based on pilot subjects, were imposed to discourage any individual from either quickly scanning the texts, or taking an inordinately long time. All subjects finished reading their texts in the given time period for their group. Subjects in the low-explanatory and standard groups tended to rate the study time more favorably than subjects in the explantory text condition, but this difference was not significant.

At the completion of the study period, subjects were given the same interpolated task used in the protocol study. Upon completion of this task, they were given a test consisting of half of the problems. There was one problem on a page with space to do any work which the subject felt might help in the solution. Subjects were given as much time as they required to complete the test; most required about 30 minutes. Upon completing the problems, subjects were asked to return 48 hours later. On the second test day, subjects were given the second form of the text, consisting of the other 12 problems. The order of the presentation of the test forms was counterbalanced.

Following the second performance test, subjects were given as much time as they required to respond to aptitude tests and to a questionnaire. These measures are described in Myers et al., (1982), where the result of their analyses is also reported.

Results

We have chosen to focus on those problems testing knowledge of Equations 1, 3, and 4; the problems within these sets revealed the widest variation in proportions of correct responses. In considering answers to these problems, we ask: why is there marked variation in proportion correct to problems requiring the same equation for solution? And why is this pattern of variation affected by text condition?

The low-explanatory and standard groups performed in similar fashion on all problems. Therefore, in the interests of simplifying the presentation, we have treated them as a single condition which we have labelled "nonexplanatory."

We begin our consideration of the results by briefly noting Table 2 which

---

Insert Table 2 about here

---

presents proportions of correct responses for each condition for each of the four problems for each of the rules presented in Table 1. It is clear from Table 2 that there was considerable variability within sets of problems, and that the pattern of proportion correct often differs markedly for the two text conditions. In what follows, we will attempt to demonstrate that the problem-specific effects are due to elements of the problem presentation which affect the ease of classification and translation, and that subjects in the nonexplanatory condition were more sensitive to these elements.

Equation 1. These four problems provide an example of our thesis that subjects in the nonexplanatory condition are sensitive to aspects of the problem statement which have little effect upon the performance of subjects

in the explanatory condition. As indicated by Table 2, proportions of correct responses in the nonexplanatory group varied greatly across the four problems testing knowledge of Equation 1, while performances of those in the explanatory group are not only much better but also quite stable.

One reason for the variation in performance in the nonexplanatory condition is that--for these subjects--some problems are much more difficult to classify than others. There are at least two aspects of a problem's statement that can cause such difficulties in misclassification: the presence of irrelevant information and the absence of key words associated with use of a particular equation. Both are illustrated by errors made in response to Problem 1B. This problem states that:

> "In the National Women's Ping-Pong League, Los Angeles has a .3 chance of winning, San Francisco has a .2 chance, San Diego has a .1 chance, New York has a .1 chance, and Philadelphia has a .3 chance. What is the probability that a West Coast team will win?"

Note that this problem contains irrelevant information; the stated probabilities for New York and Philadelphia are unnecessary. Subjects in the nonexplanatory condition attempted to use these probabilities; five subjects subtracted .1 + .3 from the correct sum (or from the product of the West Coast probabilities), four others multiplied all five probabilities, and two others divided by .1 + .3. Only one subject in the explanatory condition attempted to incorporate the irrelevant values into his answer. Apparently, subjects who do not understand the rationale for using the equations attempt to use any information presented to them. In this example, the subjects cited misclassified the problem, using Equations 2, 4, or 5 in an attempt to incorporate all the values stated.

Eight of the 32 subjects in the nonexplanatory condition used Equation 3 rather than Equation 1, multiplying probabilities instead of adding them. It is revealing that the only other of the four problems in which this type of error was frequent (nine of 12 subjects who made errors) was problem 1D, which states:

> "In a large city hospital in 1978, the proportion of births involving twins was .06; for triplets, the proportion was .02; for more than triplets, it was .01. If we define a multiple birth as one involving more than a single child, what was the probability of a multiple birth in that hospital?"

(1B) and (1D) are the only two problems in this set in which the problem does not specifically request the probability of something or something. It is also important to note that this misclassification error occurred only once for Problem (1B) and once for Problem (1D) in the explanatory condition. Apparently, the absence of the key word causes difficulty for subjects in the nonexplanatory condition.

Even when subjects in the nonexplanatory condition used the correct equation, they may have had difficulty translating the story; that is, correctly inserting values from the story into the equation. This was a major source of difficulty for the nonexplanatory group in Problem 1A. The problem states:

> "A marble is drawn from a jar containing 10 red, 30 white, 20 blue, and 15 orange marbles. What is the probability of drawing a red or white marble?"

Five subjects misclassified the problem, attempting to incorporate frequencies of orange and blue marbles into their answer. Most of the errors, however,

reveal a lack of understanding of the relation between frequency and probability. Subjects often failed to divide the sum of the numbers of red and white marbles by the total number of marbles, or divided by 100. Note that this problem was the most difficult for the nonexplanatory group but the easiest for the explanatory group who had the relation between frequency and probability explained in their text. This provides evidence of the effects of key words, irrelevant information, and translation requirements on the performance of subjects in the nonexplanatory group. The problem states:

> "The probability is .2 that Jimmy will win a race and .3 that he will finish second. What is the probability that he will win or finish second?"

Note that this problem contains no irrelevant values, includes the key word "or", and involves a simple translation of the stated values. This problem evoked the best performance in the nonexplanatory group.

Equation 3. The explanatory group did not perform as well on these problems as on those testing knowledge of Equation 1. We suspect that the rationale for multiplication of probabilities is more difficult to understand than that for adding probabilities.

In one respect, performances on this problem set were similar to those on the first problem set: proportion correct was fairly stable over problems for the explanatory group and quite favorable for the nonexplanatory group. As with the first problem set, it appears that subjects in the nonexplanatory condition are more sensitive to the statement of the problem than are subjects in the explanatory condition.

Much of the variability exhibited by the nonexplanatory group can again be traced to variations in the problems with respect to irrelevant information

and translation requirements. Problem 3D presents values that are redundant:

> "In a certain class, there are 80% men and 20% women. Also, 60% of the class are freshmen and 40% are sophomores. If sex and year are independent, what is the probability that an individual is female and a freshman?"

Four subjects in the nonexplanatory group used the wrong equation in an attempt to incorporate all four values in their answer (e.g., ".2 x .4 x .6 x .8") and one made what we would characterize as a translation error, multiplying .2 by .4. In contrast, no subjects in the explanatory group used numbers other than .2 and .6 in their answers.

The answers to Problem 3A indicate a serious translation difficulty for subjects in both conditions. The problem states that:

> "The probability that a man will be alive in 20 years is .4. The probability his wife will be alive is .6. Assuming that these are independent events, what is the probability that he will be alive and she will be dead in 20 years?"

Of 23 subjects in the nonexplanatory condition who made an error in answering this question, 11 made the response ".4 x .6;" one subject in the explanatory condition gave this answer. Nine other subjects in the nonexplanatory group substituted .4 and .6 into the wrong equation; five subjects in the explanatory group did this. It is not clear why subjects had so much difficulty classifying this problem. It is clear, however, that 20 of 32 subjects in the non-explanatory group, and six of sixteen in the explanatory group, failed to translate the stated probability of the woman being alive into the probability that she would be dead. We will shortly consider other evidence that translation of a stated probability into its complement is more difficult,

and more often overlooked in the nonexplanatory condition.

In contrast to performances on Problem 3A, the nonexplanatory group did very well in answering Problem 3B. This problem states:

> ".7 of the members of the state legislature favor a particular bill. .6 of the members are Democrats. If party and vote are independent, what is the probability that the next legislator you meet is a Democrat and votes for the bill?"

Why was this problem so easy for the subjects in the nonexplanatory group? We suggest that they have memorized the rule: "Whenever the probabilities of something and something else are required, multiply their individual probabilities." In Problem 3B, the word "and" is present and there are only two possible values to multiply. Fortunately, they are the right two values.

Equation 4. This equation is the general case of Equation 3 and, therefore, it would seem that subjects should score about the same on this problem set as on the earlier one. That was the case for subjects in the explanatory group; in the nonexplanatory group, however, the average proportion correct was .22 lower for the problems testing knowledge of Equation 4 than for those testing knowledge of Equation 3. For the explanatory condition, the proportions correct were, as in the other problem sets, relatively stable. Also, as before, the nonexplanatory group exhibits greater variability of performance over the four problems.

That the nonexplanatory group had greater difficulty with these four problems than with the third problem set may reflect two factors. First, these problems typically omit explicit use of the key word "and"; this may cause classification problems. Second, the probability of the complementary

event is often not explicitly stated; thus, there may be translation difficulties. For example, consider the wording of Problem 4A.

> "A baseball player gets a hit in 30% of his times at bat. What is the probability that he goes hitless in all four times at bat on a particular day, assuming that the result of one at bat has no effect on the result of any other at bat?"

Problems 4C and 4D are similar in that they lack the key word "and" and the student must supply one probability by subtracting the value given from 1.0.

In both conditions, there are errors of misclassification: for example, averaging over problems 4A, 4C, and 4D, five of 32 subjects in the non-explanatory condition added instead of multiplying values, and 2.33 of 16 did this in the explanatory group. There were also attempts to subtract and divide probabilities. More striking, at least in the nonexplanatory condition, were errors of translation. Averaging over the three problems, 11.67 subjects in the nonexplanatory group made errors of translation; for example, giving .34 as an answer to Problem 4A. Only 2.33 subjects made such errors in the explanatory group.

The nonexplanatory group performed reasonably well on Problem 4B which states:

> "A hockey team wins with probability of .5, loses with probability of .3 and and ties with probability of .2. What is the probability that in the next 4 games, the team wins, then ties, then wins, then loses?"

The statement of this problem seems to more closely parallel the statement of Equation 4 than does that of the other problems in this set, thus reducing the likelihood of misclassification. Furthermore, all the required probabilities are stated; translation difficulties are presumably minimal.

GENERAL DISCUSSION

In order to examine more closely the effects of an explanatory text on subsequent problem solving performance, we have recorded interviews with students as they worked on individual problems. In addition, we have analyzed the types of errors made by subjects under different text conditions to determine whether these errors would form any recognizable pattern. These interviews and error analyses provided some insights into both the successful and unsuccessful strategies our subjects used, and uncovered several common sources of difficulty underlying their attempts to solve elementary probability problems.

The protocol analyses revealed that our subjects relied primarily on critical features, or "key words," in the problem statement in order to categorize the problem. They were also able to categorize some problems on the basis of an analogy to the example provided in the text. Both the interviews and the error analyses revealed that the key word stragegy is only partially successful in categorizing probability problems because it does not always unambiguously differentiate between equations and leaves the subject at a loss when the key word is not explicit in the problem statement. In addition, even if the problem is correctly categorized, the key words are of little or no help in translating the values in the problem into the variables of the appropriate equation. Finally, subjects' protocols indicated frequent confusion about which key word signified which operation. This may be the result of the words "and," "or," and "given" having different connotations in the context of probability problems than in their more common, everyday usage with which the subjects are more familiar. For example, subjects often wrote "P(A + B) =" when the problem required the probability of some-

thing and something else. Although it is not difficult to understand these errors (it is not considered improper to say "two and two equal four"), this connotation of "and" makes it a less discriminable cue for multiplication.

On the other hand, subjects were more likely to arrive at the correct solution when they categorized problems by analogy to the example presented in the text. This sort of analogical reasoning seems less prone to the pitfalls we associated with the use of key words and thus desirable to encourage. An important factor in developing this approach to categorization may be an emphasis on explanation within texts. Although the subjects taught by an explanatory text in the protocol study did employ key words in solving problems, they are much less dependent upon those words than subjects taught by texts that place greater emphasis on rote learning of formulas. This was amply demonstrated in our analyses of errors made by subjects from two text conditions. Subjects in the nonexplanatory text condition made more classification errors on problems in which the key word was missing from the problem statement than on problems in which the key word was explicitly stated. For subjects in the explanatory text condition, the proportion correct was about the same regardless of the presence of key words; that is, their performance was relatively stable across problems testing knowledge of the same equation.

One limitation of our explanatory text may be that the text examples were different from those in the text. Several subjects indicated that they understood the text but found the problems ~~in the text~~ more "abstract" than the examples in the text. This may be because the examples in the text used countable sets (e.g., the number of marbles of a certain color, or the number of cards of a certain value) while those in the test most often

presented proportions and percentages. This observation points up the need for more research on the role of examples in the text. Based on subjects' comments about the need to have test problems more like text examples, we suspect that it will be important to identify the dimensions along which problems can vary. Our results suggest some of these dimensions: the presence or absence of countable events, of key words, of irrelevant information, and of quantities that must be transformed in order to solve the problem. If examples are to serve to illuminate concepts, and to provide an adequate base for analogical problem solving, those used in instruction should vary along the important dimensions. Such variability, rather than the sheer number of examples in the text, may well be the critical factor.

The protocol analyses also revealed that our subjects seldom used Venn and branch diagrams in solving text problems, although subjects' comments indicated that these diagrams did help them understand the accompanying examples in the text. Attempting to represent the problem statement in a diagrammatic form proved to be a source of considerable difficulty for the subjects we interviewed. These findings indicate that understanding of the representation of an example in the text does not necessarily translate into the ability to represent a new problem. To the extent that such an ability is important, practice at representing problems is a critical part of instruction in probability that is generally overlooked. How to implement such practice is an important question to be researched. It is not clear that the representations we have used, which are those most typically found in probability texts, will function best. Recent texts have incorporated more concrete diagrammatic aids, for example, pictures of individual cards and alternative outcomes of throws of dice (Freedman, Pisani, and Purves,

1978; Pagano, 1981). Systematic research on the effectiveness of various modes of representing probability problems seems in order.

The error analyses revealed that variations in test performance on problems requiring the same equation for solution were due to various elements of the problem presentation which affect the ease of classification and translation. Subjects in the nonexplanatory text condition were more sensitive to those elements than subjects in the explanatory text condition.

For subjects in th nonexplanatory text condition, the presence of irrelevant information (values not required for solution) caused errors. These subjects tended to incorporate the extra values in their answers more frequently than subjects in the explanatory text condition, which indicates that they may have a very limited understanding of the rationale underlying the use of each formula. If this is the case, it appears that the number of variables in the problem, rather than what the problem actually asks for, will dictate which formula is retrieved and used for solution. In addition, the types of errors made on certain problems ~~containing additional information~~ revealed that many subjects in the nonexplanatory text condition lacked the fundamental concept of probability as a relative frequency, even though they could correctly apply formulas to simpler problems.

The error analyses revealed two types of translation errors both of which occurred more frequently in the nonexplanatory condition. The first type was revealed by problems which required the subtraction of one of the given values from one. The most frequent error on these problems involved inserting the stated value, rather than its complement, into the equation. The second type involved the translation of a verbal statement in the problem into a probability required for the equation. These errors occurred on

problems which involved countable events. For example, on one problem testing Equation 2, several subjects translated "the probability that a randomly selected person has a birthday on the first of any month" into the value 1/12 in their answer. It is clear that there is some counting process underlying this translation, and that subjects who make this kind of error do have some sense of probability as a relative frequency. Nevertheless, they fail to take into account the entire sample space in their answer.

Research on translation skills and various sources of translation errors (e.g., Clement, Lockhead, & Monk, 1979; Paige and Simon, 1966) has indicated that translation of story problems into mathematical equations is a very common source of difficulty for many students. Although translation skills are essential to understanding and successful problem solving, they are generally taken for granted in instruction. Our analyses suggest that practice at translating probability problems should be an important part of instruction in probability. It is clear that a major factor in developing translation skills is an emphasis on explanation of the basic concepts required for successful translation over a wide range of problems.

The explanatory text used in both this and the previous study defined probability as a relative frequency of events, and presented examples which emphasized the relation between answers and counts of the numbers of ways that various events could occur. Myers et al. (1982) demonstrated that this emphasis on explanation did facilitate the translation of story problems. They used both a strict and lenient scoring criterion in analyzing story problem data. The strict criterion required that the solution be correct in all respects, while the lenient criterion accepted as correct the appropriate equation, regardless of whether the values were correctly inserted.

The difference between the two scores thus provided an index of translation errors. The explanatory group had fewer translation errors; strict and lenient criterion scores differed much less in the explanatory text condition than in the nonexplanatory text condition (See Myers, et al., 1982, for a more detailed discussion of the contrast between text conditions with respect to the strict-lenient difference).

The findings of the protocol study and the error analyses reported here complement the results of studies using similar methods to investigate the processes underlying students' problem solving performance. Chi Feltovich, & Glasser (1981) investigated the categorization of physics problems by experts and novices. Like our subjects in the nonexplanatory text condition, the novices they interviewed relied primarily on the surface features or key words in the problems, whereas the experts tended to categorize the problems according to the underlying physics principles. Although all of our subjects were novices, those in the explanatory text condition differed from the others in their processing of story problems in ways that are similar to those in which experts differ from novices.

Clement (1979) found that engineering students were particularly likely to apply the wrong formula to elementary physics problems containing extra information. Moreover, his analyses of student problem solving protocols indicated that a student's understanding of elementary principles of physics can be very weak, even though the student can remember and manipulate relevant formulas. Similarly, our subjects in the nonexplanatory text condition made more translation and classification errors than the others, even though there was some indication that subjects who read the nonexplanatory text remembered the formulas better.[2]

The clinical interview method, and the analyses of the types of errors made on paper and pencil tests, have shown that the stragegies used and the errors made are seldom unique to a single individual; rather, those strategies and errors form specifiable patterns across subjects which reflect the degree of their understanding of the concepts under study. These methods are increasingly being used as research tools in investigations of human problem solving performance. In the present study, they have brought us closer to a precise definition of what it means to understand elementary probability and to a specification of the processes such understanding might affect.

Footnotes

The research reported in the paper was supported by a grant from the National Institute of Education, NIE G80-0126, and by a grant from the National Science Foundation, BNS 80-01181. Reprint requests should be addressed to: Dr. Jerome L. Myers, Department of Psychology, University of Massachusetts, Amherst, Massachusetts, 01003.

[1]The problem states: "50% of a city's population is under 40 years of age. 60% of the city's population is male. 20% is under 40 and male. What is the probability that the next person I meet will be under 40 or male?"

[2]Myers, et al. (1982) found that subjects in the nonexplanatory text condition performed better than subjects in the explanatory condition on formula problems such as "P(A) = .4 and P(B) = .3. If A and B are incompatible events, what is P(A or B)?", but this difference was not significant. However, they also reported a follow-up study in which subjects in the nonexplanatory condition showed significantly better recall for formulas.

References

Bourne, L. E., Dominowski, R. L., & Loftus, E. F. Cognitive processes. Englewood Cliffs, N.J.: Prentice Hall, 1979.

Brooks, L. 'Nonanalytic concept formation and memory for instances.' In E. Rosch and B. B. Lloyd (Eds.), Cognition and categorization. Hinsdale, N.J.: Erlbaum, 1978.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. Categorization and representation of physics problems by experts and novices. Technical Report No. 4. University of Pittsburgh: Learning Research and Development Center, May, 1981.

Clement, J. Limitations of formula-centered approaches to problem solving in physics and engineering. Technical Report. University of Massachusetts: Department of Physics and Astronomy, July, 1979.

Clement, J., Lochhead, J., & Monk, G. Translation difficulties in learning mathematics. Technical Report. University of Massachusetts: Department of Physics and Astronomy, September, 1979.

Freedman, D., Pisani, R., & Purves, R. Statistics. New York: W. W. Norton, 1978.

Mayer, R. E. 'Acquisition processes and resilience under varying testing conditions for structurally different problem-solving procedures'. Journal of Educational Psychology, 1974, 66, 644-656.

Mayer, R. E., & Greeno, J. G. 'Structural differences between learning outcomes produced by different instructional methods'. Journal of Educational Psychology, 1972, 63, 165-173.

Mayer, R. E., Stiehl, C. C., & Greeno, J. G. 'Acquisition of understanding and skill in relation to subjects' preparation and meaningfulness of instruction'. Journal of Educational Psychology, 1975, 67, 331-350.

Myers, J. L., Hansen, R. S., Robson, R. C., & McCann, J. 'The role of explanation in learning elementary probability'. Journal of Educational Psychology. (in press).

Pagano, R. R. Understanding statistics in the behavioral sciences. St. Paul, MN: West, 1981.

Paige, J., & Simon, H. 'Cognitive processes in solving algebra word problems'. In B. Kleinmuntz (Ed.), Problem solving research, method and theory. New York: Wiley, 1968.

Table 1

Formulas Presented in the Three Texts

(1) P(A or B or C) = P(A) + P(B) + P(C)

(A, B, and C are mutually exclusive)

(2) P(A or B) = P(A) + P(B) - P(A and B)

(general case)

(3) P(A and B and C) = P(A) x P(B) x P(C)

(A, B, and C are independent)

(4) P(A, B, B, B, A) = P(A) x P(B) x P(B) x P(B) x P(A)

(5) P(A given B) = P(A and B)/P(B)

(6) P(A and B) = P(A and B) x P(B)

Table 2

Proportion Correct for Each Group for Each Problem

RULE	PROBLEM	NONEXPLANATORY (n = 32) P(C)	EXPLANATORY (n = 16) P(C)	TOTALS (n = 48) P(C)
1	A	.4063	.9375	.5833
	B	.4375	.7500	.5417
	C	.7188	.8125	.7500
	D	.6563	.8125	.7083
	TOTALS	.5547	.8281	.6458
2	A	.2188	.1250	.1875
	B	.1563	.4375	.2500
	C	.2500	.3125	.2708
	D	.1563	.2500	.1875
	TOTALS	.1953	.2813	.2240
3	A	.2813	.5000	.3542
	B	.8438	.6250	.7708
	C	.5625	.5625	.8438
	D	.5000	.6250	.5417
	TOTALS	.5469	.5781	.6276
4	A	.1563	.5000	.2708
	B	.5625	.5000	.5417
	C	.1875	.4375	.2708
	D	.3125	.4375	.3542
	TOTALS	.3047	.4688	.3594
5	A	.1875	.4375	.2708
	B	.3125	.3750	.3333
	C	.0625	.1875	.1042
	D	.1875	.4375	.2708
	TOTALS	.1875	.3594	.2448
6	A	.5938	.7500	.6458
	B	.5938	.4375	.5417
	C	.5625	.6875	.6042
	D	.6563	.6250	.6458
	TOTALS	.6016	.6250	.6094

APPENDIX 3

The Role of Explanation in Teaching Elementary Probability

Jerome L. Myers
Department of Psychology
University of Massachusetts
Amherst, MA, USA 01003

ABSTRACT

Subjects with no prior knowledge of probability read one of three texts; the texts varied in the degree of explanation and integration of basic concepts of elementary probability. Subjects in the high-explanatory text condition did equally well on both story and formula problems, falling below the formula performance of the other groups and above their story performance. A detailed analysis of error protocols for each story problem showed marked differences in which problems caused the most errors for the different groups and in the kinds of errors made. The high-explanatory group appeared to be less dependent upon key words, better able to filter irrelevant information, and more likely to insert values correctly into formulas. Thus, even with novices, qualitatively different knowledge structures can be produced by texts which vary in their emphasis on understanding.

The Role of Explanation in Teaching Elementary Probability

Jerome L. Myers
Department of Psychology
University of Massachusetts
Amherst, MA, USA 01003

Parents, educators, and mathematicians have been involved in recurrent debates about the relative merits of understanding mathematical concepts and rote memorizing computational formulas. In principle, such conflicts should be resolvable through research involving manipulation of the degree to which students understand mathematical material and observation of their performance on tests of mathematical knowledge. In actuality, we encounter difficulties because we lack both a precise definition of mathematical understanding and a specification of the processes such understanding might affect.

We have begun a research program to investigate the role of understanding in learning elementary probability by university undergraduates. Although we still lack a precise definition of understanding, we have identified three dimensions which have provided the basis for constructing short texts. The first such dimension, external connectedness, corresponds to the degree to which a text explicitly relates the concepts and formulas to be learned to the real-world knowledge the student presumably brings to the situation. The second dimension might be termed "internal connectedness," the degree to which relations among formulas and concepts are explicitly developed. For example, an equation to be learned may, or may not, be presented as a special case of a formula presented earlier in the text. The third dimension is explanation, the degree to which a rationale is provided for a particular formula. This refers not to derivations, but to intuitive justifications for procedures taught. For example one of our texts uses branch diagrams to explain why the probability of the joint occurrence of two independent events is a product of the marginal probabilities.

We will describe the texts constructed within this framework more fully in a later section. First, however, we wish to consider certain processes which should be involved in solving probability problems, and the ways in which these may be affected by the degree of understanding imparted by the text. We assume three stages: categorization of the problem, retrieval of the formula appropriate to that category, and translation of the problem--substitution of values from the statement of the problem into the retrieved formula. In general, understanding should aid both the categorization and translation of problems. Understanding may aid or hinder retrieval of formulas from memory. On the one hand, the linkage of ideas and formulas make the formulas easier to learn and recall; on the other hand, the additional explanatory and integrative material could detract from time and effort that should be spent learning the equations.

From our perspective, it makes little sense to examine total test scores as a function of the degree of understanding incorporated into the text used for instruction. Which text is better may well depend upon which processes are most important in the solution to a particular problem. For example, assume that texts which empahsize rote learning of formulas at the cost of deemphasizing explanation do provide an advantage in memorizing formulas. Then, problems in which categorization and translation are not difficult may be more easily solved by students who have less understanding of the material. On the other hand, assume that a problem provides no key words which immediately signal the problem's category, or contains considerable irrelevant information which make categorization or translation difficult; a text which has emphasized understanding may provide a definite advantage here.

In view of these considerations, we have attempted a more molecular analysis of test performance than is usually the case. For each of several text conditions, we have examined variations in performance as a function

of the type of problem. At a rather gross level, we have formula problems, which merely presented certain values (e.g. "P(A) = .6, P(B) = .4") and necessary conditions ("A and B are independent events"), and required the subject to set up the correct results (e.g. "What is P(A and B)?" requires .6 x .4 as the answer); and story problems which were considerably less transparent. Within the set of story problems, performance has been related to such variables as the presence or absence of key words (e.g. "and" implies the multiplication of probabilities), the presence of irrelevant values, and translation requirements (e.g. whether it was necessary to subtract a stated probability from one in order to solve).

Our perspective has also led us to require subjects to report their answers as operations on numbers; for example, a correct answer would be .4 x .2, not .08. This approach also provided more detailed knowledge of the types of errors made.

To summarize, we view understanding of mathematical material as a function of (1) connections of text concepts and formulas to real-world referrents; (2) integration of concepts and formulas within the text; and (3) explantion of formulas. This view has provided a basis for constructing three written treatments of elementary probability, presumably varying in the degree to which they convey understanding. Our view of the processes involved in solving problems has led us to use both formula and story problems, and to emphasize analyses of error protocols.

Method

Subjects

Forty-eight volunteers from an Introductory Psychology course at the University of Massachusetts were randomly assigned to three text conditions. None had previous exposure to probability or statistics.

Texts

There were three texts, each judged to be representative of actual chapters on probability in various introductory statistics textbooks. All

contained the six formulas shown in Table 1, with one numerical example illustrating the application of each formula. All three texts began with a brief

Table 1

Formulas Presented in the Three Texts

(1) P(A or B or C) = P(A) + P(B) + P(C)
(A, B, and C are mutually exclusive)

(2) P(A or B) = P(A) + P(B) - P(A and B)
(general case)

(3) P(A and B and C) = P(A) x P(B) x P(C)
(A, B, and C are independent)

(4) P(A, B, B, B, A) = P(A) x P(B) x P(B) x P(B) x P(A)

(5) P(A given B) = P(A and B)/P(B)

(6) P(A and B) = P(A given B) x P(B)

introduction indicating that the text contained a series of examples of probability calculations, and that each example would be followed by a formula which could be used to solve problems of that type. All three texts then contained a section labelled "What are the chances?" which introduced the concept of probability and its limiting values. This section noted that an action such as tossing a coin or observing the day's weather could have several possible outcomes, that a probability value could be assigned to these outcomes. and that these values ranged from zero to one.

Despite these similarities, the texts differed in several ways. A <u>high-explanatory</u> text defined probability as a relative frequency of events, and examples emphasized the relation between answers and counts of the numbers of ways that various events could occur. A <u>standard</u> text treated probability as a measure with certain well-defined properties (e.g. limits of one and zero) and used familiar real-world examples (e.g. the probability of rain tomorrow), but did not explain probability in terms of counting processes. A <u>low-explanatory</u> text was more abstract than either of the first two; aside

from an initial reference to dice-throwing, examples were stated in terms of such quantities as P(A) and P(B).

The texts also varied with respect to the degree of linkage among the equations presented. Only the high-explanatory text showed that one equation was a special case of another, or that one equation could be derived from another. For example, this text pointed out that the probability of the union of two mutually exclusive events (Equation 1) was a special case of the probability of the union of any two events (Equation 2).

Finally, the texts varied in the degree of explanation provided for each equation. In the high-explanatory text, Venn and branch diagrams were used to represent the information in the examples, and the equations applied in these examples were explained in terms of these diagrams. Neither of the other texts provided such explanations. For example, although all three texts stated that the probability of the joint occurrence of two independent events was the product of their probabilities and provided an example of this, only the high-explanatory text attempted to demonstrate why probabilities were multiplied in such cases.

In summary, all texts contained an example followed by the equation, as well as sufficient definitions of terms such as mutual exclusivity and independence. In the high-explanatory text, the relation of probability to counting was emphasized, pictorial aids were used, and the equation was presented as a natural development ensuing from the example.

Performance Tests

Two tests were constructed, each containing two sets of six story problems and two sets of six formula problems. Within each set, there was one problem testing each of the six equations in the text. The order of sets within a test was counterbalanced over subjects so that there were four sequences, with the constraint that story and formula sets would alternate. Half of the subjects received one test first and the remaining subjects received the other test first.

Formula and story problems differed in the degree to which the given information was explicit in the question. For example, a formula problem testing knowledge of Equation 1 would state "P(A) = .4 and P(B) = .3. If A and B are incompatible, what is the value of P(A or B)?" In contrast, a story problem testing the same knowledge would read as follows: "The probability is .20 that Jimmy will win a race and .30 that he will finish second. What is the probability that he will win or finish second?"

Procedure

Subjects were tested individually. They were instructed to read the text booklets silently and at their own rate, trying to understand the explanations and examples. The standard and low-explanatory groups, whose texts were four pages in length, were told that they would have 15 minutes to finish; the high-explanatory group, whose text was 14 pages in length, was given 25 minutes. These limits, based on pilot subjects, were imposed to discourage any individual from either quickly scanning the texts, or taking an inordinately long time. All subjects finished reading their texts in the given time period for their group. Subjects in the low-explanatory and standard groups tended to rate the study time more favorably than subjects in the high-explanatory text condition, but this difference was not significant.

At the completion of the study period, subjects were given two minutes to study a list of 15 words unrelated to probability and were then required to write down as many words as they could recall on a blank sheet of paper.

Upon completion of the interpolated task, subjects were given the test. There was one problem on a page with space to do any work which the subject felt might help in the solution. Subjects were given as much time as they required to complete the test; most required about 30 minutes. Upon completing the problems, subjects were asked to return 48 hours later. On the second test day, subjects were given the second form of the test. The order of presentation of the test forms was counterbalanced.

Results and Discussion

Performance test data were scored according to both a strict and a lenient criterion. The strict criterion required that the subject set up the answer correctly in all respects; the lenient criterion gave credit if the subject used the correct equation but, for example, inserted the wrong values. This distinction applies only to performances on story problems; strict and lenient scores were the same for formula problems.

A Summary of Test Data

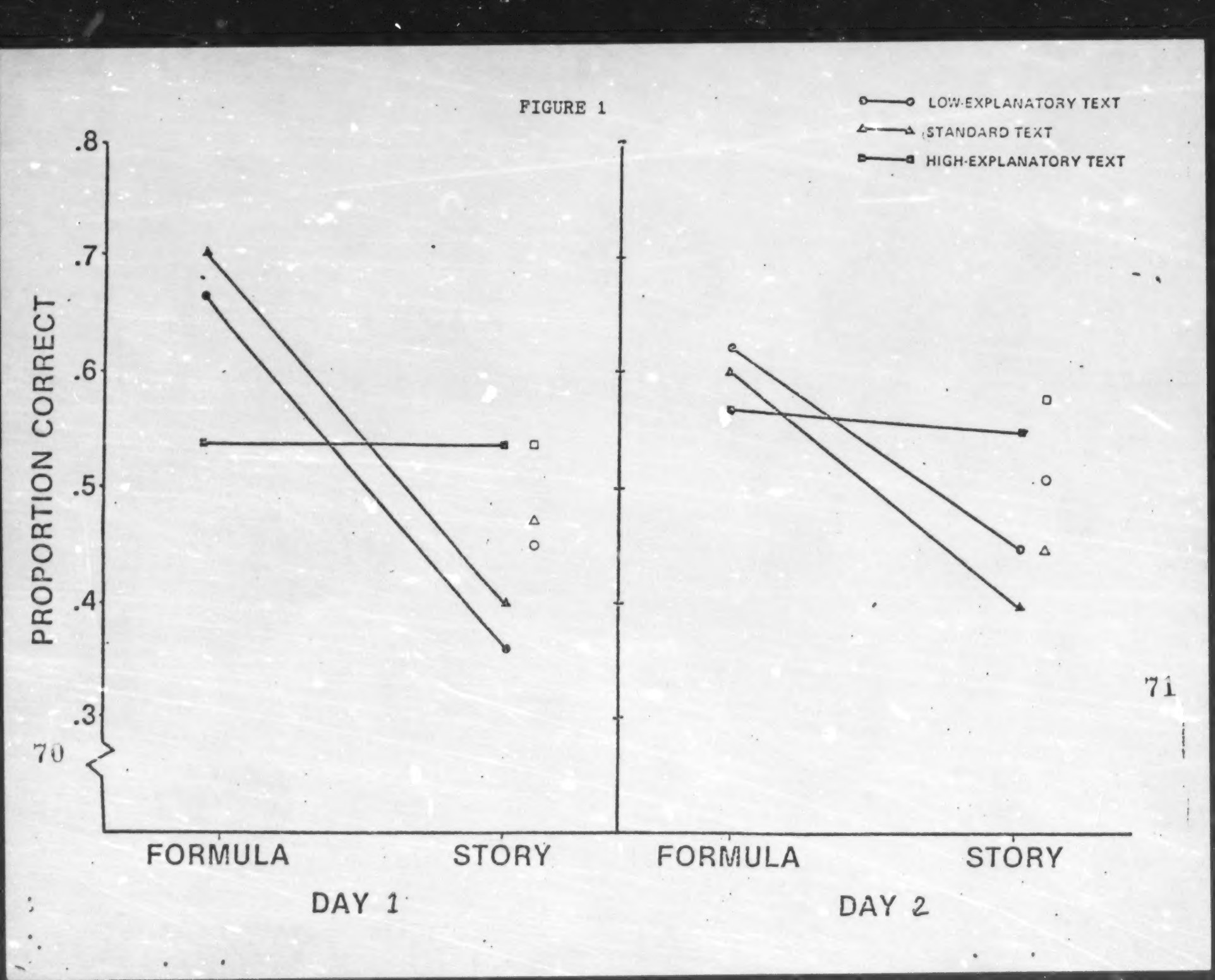
Figure 1 (next page) summarizes the data. The points on the lines are for the strict criterion; the disconnected open symbols are for the lenient criterion. There are two important observations to be made about the means. First, considering the strict-criterion means, the relative performances of the three text conditions strongly depend upon the type of problem ($p < .001$). The two nonexplanatory groups appear to know the formulas better than the high-explanatory group but they have more difficulty in applying that knowledge to story problems.

The second point to note in Figure 1 is that strict-criterion scores are significantly lower than lenient-criterion scores for the two nonexplanatory groups ($p < .001$). Apparently, subjects in these two groups may correctly categorize the problem and recall the appropriate formula but are not able to translate the values in the problem statement. In contrast, when high-explanatory subjects know what operations are to be performed to solve a problem, they usually perform them on the correct set of values; there is little difference between their strict and lenient-criterion scores.

A More Detailed Analysis of Individual Problems

It is impractical to examine the numbers and types of errors made to each of the 24 story problems for each of the three text conditions. Therefore, we will focus on those problems testing knowledge of Equations 1, 3, and 4: the problems within these sets reveal the widest variation in proportions

FIGURE 1


LOW-EXPLANATORY TEXT
STANDARD TEXT
HIGH-EXPLANATORY TEXT
PROPORTION CORRECT
.8
.7
.6
.5
.4
.3
FORMULA
STORY
FORMULA
STORY
DAY 1
DAY 2

of correct responses. In considering these problems, we ask: Why is there marked variation in proportion correct to problems requiring the same equation for solution? And why is this pattern of variation affected by text condition?

The low-explanatory and standard groups performed in similar fashion on all problems. Therefore, in the interests of simplifying the presentation, we have treated them as a single condition which we have labeled "nonexplanatory." We will refer to the high-explanatory group as "explanatory."

Equation 1. Performances on these four problems indicate that subjects in the nonexplanatory condition are sensitive to aspects of the problem statement which have little effect upon the performance of subjects in the explanatory condition. Proportions of correct responses in the nonexplanatory group varied greatly across the four problems testing knowledge of Equation 1: (a) .375, (b) .4375, (c) .625, and (d) .7188. In contrast, the performances of those in the explanatory group are not only much better but also quite stable: (a) .9375, (b) .75, (c) .8125, and (d) .8125.

Examining the answers to (a) and (b) more closely, we find clear evidence of failures to categorize or translate these problems correctly, but only in the nonexplanatory condition. Consider problem (b):

> "In the National Women's Ping Pong League, Los Angeles has a .3 chance of winning, San Francisco has a .2 chance, San Diego has a .1 chance, New York has a .1 chance, and Philadelphia has a .3 chance. What is the probability that a west coast team will win?"

Note that this problem contains irrelevant information; the stated probabilities for New York and Philadelphia are unnecessary. Subjects in the nonexplanatory condition attempted to use these probabilities; five subjects subtracted .1 + .3 from the correct sum (or from the product of the west coast probabilities), four others multiplied all five probabilities, and two others divided by .1 + .3. Only one subject in the explanatory condition attempted to incorporate the irrelevant values into his anwer. Apparently, subjects who do not understand the rationale for using the equations attempt to use any information presented to them. In this example, the subjects

cited misclassified the problem, using Equations 2, 4, or 5 in an attempt to incorporate all the values stated.

Problem (b) also indicates that subjects in the nonexplanatory condition have difficulty when there is no key word to provide a basis for classification. Eight of the 32 subjects in this condition used Equation 3 rather than Equation 1, multiplying probabilities instead of adding them. It is revealing that the only other of the four problems in which this type of error was frequent (nine of 12 subjects who made errors) was (c); (b) and (c) are the only two problems in this set in which the problem does not specifically request the probability of something or something. It is also important to note that this misclassification error occurred only once for problem (b) and once for problem (c) in the explanatory condition.

Even when subjects in the nonexplanatory condition used the correct equation, they may have had difficulty translating the story; that is, correctly inserting values from the story into the equation. This was a major source of difficulty for the nonexplanatory group in problem (a). The problem presents frequencies of different color marbles in a jar, and then asks for the probability of drawing a red or white marble. Five subjects misclassified the problem, attempting to incorporate frequencies of orange and blue marbles into their answer. Most of the errors, however, reveal a lack of understanding of the relation between frequency and probability. Subjects often failed to divide the sum of the numbers of red and white marbles by the total number of marbles, or divided by 100. Note that this problem was the most difficult for the nonexplanatory group but the easiest for the explanatory group who had the relation between frequency and probability explained in their text.

The performance of the nonexplanatory group on problem (d) provides further evidence of the validity of our analysis of classification and translation errors. This problem evoked the best performance from the nonexplanatory

group. Not surprisingly, in view of our analysis, it states exactly two values, includes the key word "or," and requires no translation of the stated values.

Equation 3. The explanatory group did not perform as well on these problems as on those testing knowledge of Equation 1. We suspect that the rationale for multiplication of probabilities is more difficult to understand than that for adding probabilities.

In one respect, performances on this problem set were similar to those on the first problem set: proportion correct was fairly stable over problems for the explanatory group and quite variable for the nonexplanatory group. For the explanatory group, these proportions were (a) .5, (b) .5625, (c) .625, and (d) .5625; for the nonexplanatory group, the proportions were (a) .2813, (b) .5, (c) .625, and (d) .8438. As with the first problem set, it appears that subjects in the nonexplanatory condition are more sensitive to the statement of the problem than are subjects in the explanatory condition.

Much of the variability exhibited by the nonexplanatory group can again be traced to variations in the problems with respect to irrelevant information and translation requirements. Problem (b) presented values that were redundant (in essence, presenting P(A), P(not-A), P(B), P(not-B). Four subjects in the nonexplanatory group used the wrong equation in an attempt to incorporate all four values in their answer (e.g. ".2 x .4 x .6 x .8") and one made what we would characterize as a translation error, multiplying .2 by .4. In contrast, no subject in the explanatory group used numbers other than .2 and .6 in their answers.

The answers to problem (a) indicate a serious translation difficulty. The problem states P(A) and P(not-B), and asks for P(A and B). Of 23 subjects in the nonexplanatory condition who erred in response to this problem, 11 multiplied P(A) by P(B); only one subject in the explanatory condition did this.

In contrast to performances on problem (a), the nonexplanatory group did very well in answering problem (d). This states:

> ".7 of the members of the state legislature favor a particular bill. .6 of the members are Democrats. If party and vote are independent, what is the probability that the next legislator you meet is a Democrat and votes for the bill?"

We suggest that this problem was very easy for the nonexplanatory group because they have memorized the rule: "Whenever the probabilities of something and something else are required, multiply their individual probabilities." In problem (d), the word "and" is present and there are only two possible values to multiply. Fortunately, they are the right two values. The same circumstances hold for problem (a) except that the values given are not right; one value must be subtracted from 1.0 before multiplying.

Equation 4. This equation is the general case of Equation 3 and, therefore, it would seem that subjects should score about the same on this problem set as on the earlier one. That was the case for subjects in the explanatory group; in the nonexplanatory group, however, the average proportion correct was .22 lower for the problems testing knowledge of Equation 4 than for those testing knowledge of Equation 3. For the explanatory condition, the proportions correct were, as in the other problem sets, relatively stable: (a) .5, (b) .4375, (c) .625, and (d) .5625. Also, as before, the nonexplanatory group exhibits greater variability of performance over the four problems: the relevant proportions were (a) .1563, (b) .1875, (c) .3438, and (d) .625.

That the nonexplanatory group had greater difficulty with these four problems than with the third problem set may reflect two factors. First, these problems typically omit explicit use of the key word "and;" this may cause classification problems. Second, the probability of the complementary event is often not explicitly stated; thus, there may be translation difficulties.

Both conditions exhibit about 16% classification errors as indicated by use of operations other than multiplication. More striking, at least

in the nonexplanatory condition, were errors of translation indicated by using the stated probability when its complement was appropriate. Averaging over problems (a), (b), and (c), 11.67 subjects in the nonexplanatory group made such errors. Only 2.33 subjects did so in the explanatory group.

The nonexplanatory group performed reasonably well on problem (d) which states:

> "A hockey team wins with probability .5, loses with probability of .3, and ties with probability of .2. What is the probability that in the next four games, the team wins, then loses, then wins, then loses?"

The statement of this problem more closely parallels the statement of Equation 4 than does that of the other problems in this set, thus reducing the likelihood of misclassification. Furthermore, all the required probabilities are stated; translation difficulties are presumably minimal.

Conclusions

Very different patterns of knowledge appear to be present in subjects in the nonexplanatory (standard and low-explanatory texts) and explanatory conditions. Subjects in the two nonexplanatory groups performed considerably less well on story than on formula problems, and often used the correct formula for a problem (that is, met the lenient criterion) but failed to solve it. A closer look at answers to story problems revealed that subjects in the nonexplanatory conditons often required the explicit presence of key words which unambiguously pointed to certain operations, tended to misclassify problems in the presence of irrelevant or redundant information, and made many errors when the values in the story required modification before insertion into the formula. In contrast, subjects in the high-explanatory condition performed equally well on story and formula problems, tended to solve whenever they showed evidence of knowing the appropriate formula, and were considerably less hindered by absence of key words, the presence of irrelevant information, and the need to translate values in the story. The situation may be summarized by noting that, although all our subjects were novices, those in the high-

explanatory condition seem to share more of the characteristics of experts than do those in the other two conditions.

The results of more recent studies lead us to believe that the solution processes in the explanatory and nonexplanatory conditions are qualitatively different. In one study, subjects in a high-explanatory condition had only 60% correct recall of equations (as opposed to 90% for a standard group), but were able to perform correctly on 53% of story problems. This result suggests to us (as does the equivalence of formula and story scores in the study just detailed) that subjects in the high-explanatory condition often may not retrieve the formula at all. Another study in which subjects thought aloud while solving supports this conjecture; after studying a high-explanatory text, students tended to attempt to construct solutions, often using the diagrammatic aids provided by the text. Rather than recalling formulas, they attempted to recall examples from the text which they felt were similar to the problem at hand, and to use these as a model for solution.

Much remains to be done. High on the agenda must be the use of other measures of comprehension--both problems demanding more complex processing and transfer to new materials such as the Binomial and Bayes' theorems. We suspect that such research will provide further evidence of the role understanding can play in mathematical problem-solving.

Acknowledgments

This research was supported by a grant from the National Institute of Education, NIE G80-0126, and by a grant from the National Science Foundation, BNS 80-01181. Randall Hansen, Joan McCann, and Rikki Robson made major contributions to all phases of the research reported in this paper.